Project Title: Predicting Newly Acquired Customer Quality

Project Objectives:

The primary goal of this project is to measure the quality of newly acquired customers (i.e who have at least 1 successful order) for an e-commerce company using data analysis and machine learning techniques. The key objectives to achieve this goal are as follows:

1- Define the target variable that will be used to measure customer quality.
2- Predict the quality of new customers.
3- Segment customers to identify distinct customer groups by using your output.
4- Recommend strategies to improve customer loyalty.

Evaluation

We mostly consider these factors to evaluate your study:

- Your capabilities to explore and make sense of given datasets while generating meaningful and clearly explained insights.
- The amount and conclusiveness of evidence you provide to support your analysis.
- A presentation about the modeling process that is most convenient for you (pdf, pptx, xlsx, ipynb etc. are all welcomed).
- Your code (Clean and commented code is a plus).
- Your model's predictions for given clients.

Data Description:

market_order: Users' basket purchase data

client_id: The unique identifier assigned to a user when they register on the app. This ID is used to track the user's activities and purchases on the platform.
order_id: The unique identifier of the particular order.
checkoutdate: The date when the user checked out or finalized their purchase.
createdate: The date when the basket was created or when the first product was added to it.
city_id: The unique identifier for the city of the user.
devicename: The name or model of the device used by the client to make the purchase.
addressname: Name or label associated with the delivery address (e.g., "Home," "Work").
distance: The distance between the warehouse and the delivery address.
aandm: Financial metrics associated with advertising and marketing costs.
basketvalue: The total value of the products in the basket.
chargedamount: The total amount charged to the user including all fees and discounts.
deliveryfee: Fee charged for delivery.

discountamount: Amount of discount given on the basket.
suppliersupport: Financial support provided by the supplier, if any.
thirdpartysupport: Financial support provided by third parties, if applicable.
ratedate: The date when the user provided a rating for their purchase.
rateisskipped: Indicates if the user skipped giving a rating.
rating: The rating given by the user for their purchase experience.
ratereason: The reason associated with the rating, usually when the rating is negative or below a certain threshold.
deliverdate: The date when the products were delivered to the user.
warehouse_id: Unique identifier for the warehouse from which products were dispatched.
hex_id: Hex ID related to the geolocation coordinates of the delivery location.

app_open: **This dataset captures details when a user opens the e-commerce application.**
client_id: The unique identifier of the user who opened the app.
appopendate: The timestamp when the user opened the app.

promo_detail: **This dataset captures details about promotions.**
order_id: The unique identifier of the particular order.
promo_id: Unique identifier for the promotion.
responsibledepartment_id: Department responsible for the promotion.
promoobjective: Objective or goal of the promotion, e.g., increase sales, attract new customers.

product_detail: This dataset captures details about products.
order_id: The unique identifier of the particular order.
_pos: Position or rank of the product, possibly in search results or listings.
count: Quantity of the product.
product_id: Unique identifier for the product.
price: Price of the product.
subcategory_id: Unique identifier for the subcategory of the product.
brand_id: Unique identifier for the brand of the product.
mastercategory_id: Unique identifier for the main or master category of the product.